局部扩展的遗传优化重叠社区发现方法 *

楚杨杰，杨忠保，洪　叶

(武汉理工大学 理学院，武汉 430070)

摘　要：重叠社区结构是复杂网络的一种重要的特征，提出了一种局部扩展的遗传优化重叠社区发现（LEGAOCD）。借鉴局部扩展的重叠社区发现方法的思想，将少数的核心节点构成模体；同时，利用了三角形模体来判断社区的稳定性度量问题，从而量化社区结构稳定性；然后通过改进的遗传优化算法策略分配它们应归属的社区；最后通过两个评价目标函数得到高质量的重叠社区结构。该算法在数据集上与经典的 CPM 算法、COPRA 算法作比较，实验结果表明，LEGAOCD 算法在检测重叠社区结构和重叠节点方面具有较优的性能。

关键词：局部扩展；遗传算法；重叠社区发现；核心节点；多目标优化

中图分类号：TP301.6　　doi: 10.3969/j.issn.1001-3695.2017.10.0934

Local extension approach through genetic algorithm for overlapping community detection

Chu Yangjie, Yang Zhongbao, Hong Ye

(*School of Science*, *Wuhan University of Technology*, *Wuhan* 430070, *China*)

Abstract : Overlapping community structure is one of the most important features of complex network. This study proposed a local extended genetic algorithm optimization overlapping community detection(LEGAOCD) . It makes a few core nodes be constructed as die body, yet regards the main idea of local extended overlapping community detection as reference; at the same time, this paper uses the triangular model to judge the stability measure of the community, so as to quantify the stability of community structure. Then, the improved strategy of genetic algorithm is used to allocate the communities where they belong. Finally, the high-quality overlapping community structure is obtained by two discriminant objective functions. After that, the LEGAOCD is compared with classical CPM and COPRA algorithms on the data sets, the results show that LEGAOCD possesses excellent comparatively in the aspects of detecting overlapping community structure and overlapping nodes.

Key Words: local extend; genetic algorithm; overlapping community detection; core node; multi-objective optimization

0　引言

社区是由网络共享相同属性的节点构成的，社区在表示的真实系统中扮演着至关重要的角色，挖掘网络中的社区结构对于帮助分析网络的属性功能具有深刻的意义[1]。目前有一些社区发现的方法旨在挖掘网络中的标准划分，即划分结果中的每个节点只属于一个社区。然而，现实生活中，节点往往会同时隶属于多个社区，如在社会网络中，每一人会有很多身份并分别隶属于不同的生活圈诸如同学圈、朋友圈、家庭或其他爱好圈。在科学家协作网络中，一名科学研究者研究不同的领域，这些领域属于不同的研科团队。在蛋白质网络中，一个氨基酸往往也具有不同的生物功能[2]。因此，重叠社区是网络中节点的集合，社区内节点同时隶属多个不同的社区，社区内部节点间的联系较为紧密，而属于不同社区的节点之间的联系较为稀疏。从而重叠社区发现则更加真实反映出网络的组成结构。

静态网络环境下，重叠社区发现算法大约分为五类，分别是派系过滤方法、局部扩展方法、线图划分方法、模糊聚类方法和基于 Agent 方法[1]。其中经典算法是基于团渗理论方法（CPM）[3]、基于智能体类方法（COPRA）[4]和社交标签传播方法（SLPA)[5]。经典的团渗透算法属于派系过滤方法，它能发现较大的社区，挖掘小社区的可能性较低；算法运行的时间和空间开销过高，使得无法应用于较大网络。COPRA 算法属于基于 Agent 方法，它随机选取标签，将会导致该算法收敛性能较差，重叠社区划分的结果不够稳定。SLPA 算法也属于基于 Agent 方法，它主要设置不同的概率准则以及调整参数对应不同的标签，导致该算法在实际应用过程中对不同网络进行参数调节，难以推广。

本文提出了一种局部扩展的遗传优化重叠社区发现（LEGAOCD）。借鉴局部扩展的重叠社区发现方法的思想，将少数的核心节点构成模体；同时，本文利用了三角形模体来判

基金项目：中央高校基本科研业务费专项资金资助（2017IB014）

作者简介：楚杨杰（1969-），男，副教授，博士，主要研究方向为复杂系统及建模；杨忠保（1988-），男，硕士研究生，主要研究方向为复杂性分析与评价（623301295@qq.com）；洪叶（1993-），女，硕士研究生，主要研究方向为复杂性分析与评价.

断社区的稳定性度量问题，从而量化社区结构稳定性。然后通过改进的遗传优化算法策略分配它们应归属的社区；最后通过两个评价目标函数得到高质量的重叠社区结构。该算法在数据集上与经典的 CPM 算法、COPRA 算法作比较，实验结果表明，LEGAOCD 算法在检测重叠社区结构和重叠节点方面具有较优的性能。

1 基本概念

假设一个网络用一个有向图三元组进行表示 $G=(V,E,W)$，其中 V 是表示图的节点集合，E 表示表示图中的边集合，W 表示图的边权集合。有权方向 $w\in(pr,nr)$ 表示相关系数，pr 为正相关系数，nr 为负相关系数。而在无权网络 $G(V,E)$ 中，如果节点 v_i 与节点 v_j 中间有边相接，则 $e=1$，否则，$e=0$，所以节点 v_i 的度 k_i 表示为 $k_i=\sum_{v_j\in V}e$，表示与节点 v_i 相连的边的数目。节点 v_i 的点权 s_i 定义为与它关联的边权之和，也称为点强度。

定义 1 重叠网络社区结构。网络社区结构就是网络节点 V 集合的一个划分 m 方案 $P=\{c_1,c_2,\cdots,c_m\}$，其中，c_i 必须满足如下条件：$c_i\subseteq V, c_i\neq\phi(i=1,2,\cdots,m)$，$\bigcup_{i=1}^{m}c_i=V$ 与 $c_i\cap c_j\neq\phi(i\neq j)$。所以社区的点强度的关系为：$s_i=(s_i^+)^{in}-(s_i^-)^{in}+(s_i^+)^{out}-(s_i^-)^{out}$。其中，

$(s_i^+)^{in}=\sum_{v_j\in c,w\in pr}w$ 表示社区 c_i 内的节点之间的正相关的内部点强度；

$(s_i^-)^{in}=\sum_{v_j\in c,w\in nr}w$ 表示社区 c_i 内的节点之间的负相关的内部点强度；

$(s_i^+)^{out}=\sum_{v_j\notin c,w\in pr}w$ 表示社区 c_i 内的节点之间的正相关的外部点强度；$(s_i^-)^{out}=\sum_{v_j\notin c,w\in nr}w$ 表示社区 c_i 内的节点之间的负相关的外部点强度。

网络中节点的重要程度不同；核心节点具有较高重要性，而普通节点则是网络中的参与者。社区是由核心节点和追随节点组成的一个社区结构，划分网络结构，要找出社区的核心节点[6]。核心节点对网络的运行具有主导的作用。

定义 2 最大核心节点的度定义。设核心节点的标签集为 $\Omega(u)=(u_1,u_2,\cdots,u_k)$，则最大核心节点的度定义为

$$ion(u_t)=\frac{\deg ree(u_t)\cdot num(u_t)}{k^2} \tag{1}$$

$$\mathrm{Np}(u)=\underset{u_t\in\Omega(u)}{\arg}\ \max(ion(u_t)) \tag{2}$$

两个公式中，如果网络节点 i 为核心节点，则 $k_i=\deg ree(u_t)$，$s_i=num(u_t)$ 分别表示重新排序后的节点度和点强度。如果这样单纯依赖排序结果，选择初始核心节点会带来潜在风险，多个核心节点可能在相同社区，按从大到小的选择，可能会把一个大的社区分解为若个小社区；为了避免这些问题，在核心节点选择上附加一些约束：为了避免核心节点与外界节点连接较少，选择的核心节点的点强度必须大于网络中所有的节点的平均点强度。为了先后选出的几个核心节点在同一社区，当前选择的核心节点必须和已有核心节点相邻的。

定义 3 核心节点共同邻居。核心节点 u_i 与节点 u_j 的共同邻居定义为

$$C_{ij}=neighbor(u_i)\cap neighbor(u_j) \tag{3}$$

其中，$neighbor(u_i)$ 表示核心节点 u_i 的邻居节点集。

定义 4 模体结构。模体介于个体与社区之间，少数几个节点连接构成，模体是社区内部成员之间基本的连接模式，揭示网络的演化规律。

定义 5 加权社区聚类。加权社区聚类[8]（Weighted Community Clustering)是测量网络拓扑性的目标函数，它是依赖于社区的三角形模体（三个节点的模体）决定该社区的稳定性，从而量化社区结构质量。加权社区聚类定义如下：

$$WCC(u,NP)=\begin{cases}\frac{t(u,NP)}{t(u,V)}\bullet\frac{vt(u,V)}{|NP\setminus u|+vt(u,\mathrm{V}\setminus\mathrm{NP})} & if\quad t(u,\mathrm{V})>0\\ 0 & otherwise\end{cases} \tag{4}$$

其中：$t(u,NP)$ 表示核心节点 u 与节点集合 NP 构成三角形模体的数目；$vt(u,V)$ 表示节点 u 与节点集合 V 至少构成一个三角形模体的节点数目；$|NP\setminus u|$ 表示节点集合 NP 除去节点 u 的数目。

定义 6 节点与社区的贴近度。根据文献[9]节点与社区的贴近度定义如下：

$$F(\mathrm{u,c})=\frac{\left(\left(s_i^+\right)^{in}-\left(s_i^-\right)^{in}\right)^{c+\{u\}}}{s_i^a}-\frac{\left(\left(s_{i-1}^+\right)^{in}-\left(s_{i-1}^-\right)^{in}\right)^{c}}{s_{i-1}^a} \tag{5}$$

而重叠节点与社区的贴近度定义如下：

$$OD(u,c)=\frac{F(u,c)}{\sum_{i=1,2,\cdots,n\wedge u\in c_i}F(u,c_i)} \tag{6}$$

其中：a 为分辨率系数，用来控制社区的大小；c 为社区，$c+\{u\}$ 是指在社区 c 中加入节点 u 形成新的社区。

定义 7：模块度。模块度用来对重叠社区划分的整体质量作出一个定量的评价，根据正负相关的边权情况，采取合适性能的模块度，根据文献[9]改进定义如下：

$$SQ=\frac{1}{2s^+-2s^-}\sum_{c\in p}\sum_{v_i,v_j\in V}\delta_{cv_i}\delta_{cv_j}\left(w_{v_iv_j}-\left(\frac{s_i^+s_j^+}{2s^+}-\frac{s_i^-s_j^-}{2s^-}\right)\right) \tag{7}$$

其中：s_i^+ 与 s_i^- 分别表示关于节点 i 所有的正与负相关的权重之和；特别地，$w_{v_iv_j}$ 表示网络相关权重的邻接矩阵；如果节点 v_i 属于社区 c 内，则 $\delta_{cv_i}=1$；否则，$\delta_{cv_i}=0$。

2 局部扩展的遗传优化重叠社区发现方法

本文以遗传优化算法为背景知识解决网络中的重叠社区发现问题，旨在挖掘出高质量的重叠社区结构。本文所提的算法流程框架如下：

算法 1： LEGAOCD 算法

输入：网络拓扑结构 $G=(V,E,W)$

输出：重叠社区划分集合 $P=\{c_1,c_2,\cdots,c_m\}$

1：为网络的核心节点进行编码；

2：根据核心节点的邻接矩阵 A 和点强度的关联矩阵 W；

3：种群的初始化；

4：$NP \leftarrow \phi; u_t \leftarrow 0 (t=1,2,\cdots,n)$

5：for each core node u in V do

6：if $WCC(u_t, NP) \neq 0$；$t \leftarrow t+1$

7：while $(n \leq k)$ do

8：if $WCC(u_{t+1}, \mathrm{NP}) \geq \mathrm{WCC}(\mathrm{u}_t, NP)$ then

9：$NP \leftarrow NP \cup \{u_{t+1}\}$

10：else

11：计算适应度函数 $F(u,c)$；选择算子；

12：进化操作，具体包括均匀交叉和变异；

13：子代种群，合并种群，精英选择，返回步骤 3；

14：当 $n=k$ 时，输出众多粗糙的重叠社区结构。

15：end

16：end

17：while $(n \leq m)$ do

18：foreach overlapping node u in V do

19：if $WCC(u, \mathrm{NP}_t) \geq \mathrm{WCC}(\mathrm{u}, NP_{t-1})$ then

20：use SQ

21：else

22：use $OD(u,\mathrm{c})$

23：当 $n=m$ 时，得到高质量的重叠社区结构

24：end

25：end

26：将所有的节点按照共有邻居有序扩展集合得到 P。

LEGAOCD 算法的第 1、2 行属于准备工作，将网络拓扑结构量化处理；第 3~16 行属于改进的遗传优化算法策略分配它们应归属的社区；第 17~26 行属于社区结构质量提高预处理阶段。

算法时间复杂度分析：n 为网络节点的数目，k 为网络中核心节点的数目，m 为重叠社区结构的数目。LEGAOCD 算法的第 1、2 行用时为 $O(n)$，第 3~16 行是属于遗传过程，算法的解码时间为 $O(k)$，选择、交叉、变异操作时间为 $O(\mathrm{m+k})$，第 17~26 行用时为 $O(m)$。在遗传算法中，种群数量为 p，迭代次数为 g。则 LEGAOCD 算法时间复杂度为 $O(gp\log(\mathrm{p}\bullet(m+k+n)))$。

2.1 个体的编码与解码

初始化，网络的所有的节点、社区、各节点与算法中的种群、个体、基因互相映射。借鉴标签传播算法的思想，标签传播算法能够产生有一定社区结构的个体，节点的点强度越大，那么这个节点对它隶属的社区稳定性和影响力也就越大。随机产生个体，多个个体构成了一个群体。本文采用字符串编码方式，网络的非重叠核心节点所属社区仅仅是一个标志符，而重叠核心节点有多个标识符。图 1 展示了一个网络划分和相应有编码。如图 1 所示的网络可能存在两个个体(1,1,12,2,2,2)和（2,2,3,3,3,3)，第一个个体中存在两个重叠社区{1,2,3}和{3,4,5,6}；第二个个体存在非重叠的两个社区{1,2}和{3,4,5,6}. 如果这两个个体之间进行交叉操作后得到结果可能是（1,1,1,1,1,1),则整个社区结构被破坏。解码过程针对任意个体，其初始社区标签为 $NP(i)(i=1,2,\cdots,k-1)$，设 $NP(1)=1$，如果 $NP(1)=NP(2)$，社区结构合并，如同算法步骤 9。如果 $NP(1) \neq NP(2)$，则 $NP(2)=2$，以此类推，如果此时 $k-1$ 时是当前最大社区标签值才结束。整个解码过程，个体(1,1,12,2,2,2)和（2,2,3,3,3,3)解码为(1,1,12,2,2,2)和(1,1,12,2,2,2)，然后再和其他个体进行交叉操作，保留稳定性较强的社区结构。解码后，节点的标签值表示节点属于社区编号。

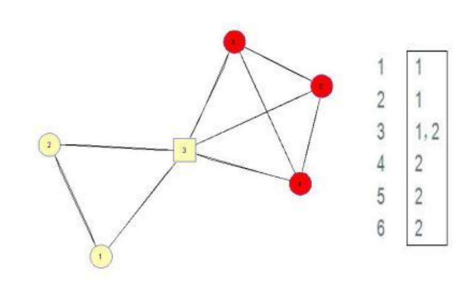


图 1　6 个节点网络被划分两个社区结构{1,2,3}和{3,4,5,6}

2.2 选择

选择的目的是为了从当前群体中选出优良的个体，使它们有机会作为父代为下一代繁殖子孙。遗传算法通过选择过程体现这一思想，进行选择的原则是适应性强的个体为下一代贡献一个或多个后代的概率大。选择体现了达尔文的适者生存原则，本文采用轮盘选择方式[11]，利用定义 6 的计算公式为适应度函数。

2.3 遗传进化操作更新策略

遗传进化操作包括交叉和变异。交叉是通过交换父代染色体中的部分基因产生新的后代，新个体组合了其父代个体的特性。交叉体现了信息交换的思想。本文采用均匀交叉的二维交叉的交叉算子[12]，将父代染色体两两分组，就会随机产生一个均匀块交叉算子，交叉算子将根据预设的交叉概率交换两个父代染色体在矩形块中的基因部分，模体、个体、社区都能用矩形块表示出来，例如，两个矩形块(1,1,12,2,2,2)和（2,2,3,3,3,3) 交叉后，解码为(1,1,12,2,2,2)和(1,1,12,2,2,2)。从而改变节点的标签值，矩形块的大小和位置是随机产生的。使用均匀块交叉算子能保证交叉产生的后代染色体的每个基因都是已有边的邻接边，使得算法仍然是朝着矩形块（模体）稳定的方向进行网络结构划分。变异在群体中随机选择一个个体，对于选中的个体以一定的概率随机地改变社区里某个节点的标签植。变异算子是以预设的概率改变个体的某个节点的标签值，变异发生的概率很低。

3 实验结果与分析

LEGAOCD 算法、CPM 算法、COPRA 算法均在 Matlab(R2010b)编码实现。实验环境设置为：windows 7 操作系统，AMD A-8 2.10GHz,500GB 内存。根据遗传算法的设置参数实验可知：标准互信息的值（NMI 的值）随着交叉概率和变异概率的变化，NMI 的值没有表现出明显的变化，设置实验参数

一般采用高的交叉概率和低的变异概率[13]。所以，为了方便检测算法的性能，实验参数设置与一种新的遗传算法重叠社区检测方法[14]相同，实验参数设为：控制参数为 $a=0.97$，交叉概率为 $p_c=0.8$，变异概率为 $p_m=0.2$，种群数量为 $p=200$，迭代次数为 $g=100$。LEGAOCD 算法、CPM 算法、COPRA 算法的结果为运行 50 次取平均值。

3.1　评价标准

F-score 是由准确率和召回率的组成计算得到的，用来衡量算法检测的重叠节点的准确性的一个指标[7]，公式可推导如下：

$$P=\frac{|L_{result}|\cap|L_{true}|}{|L_{result}|},\quad R=\frac{|L_{result}|\cap|L_{true}|}{|L_{true}|} \tag{8}$$

$$F=\frac{2P\bullet R}{P+R} \tag{9}$$

三个公式中，$|L_{result}|$ 表示算法得到的社区，$|L_{true}|$ 表示真实的社区；P 为准确率，它是指正确检测的重叠节点数目占所有检测到的重叠节点数目的比例；R 为召回率，它是指正确检测的重叠节点数目占网络中真正的重叠节点数目的比例。

扩展的规范互信息量(normalized mutual information,NMI)用来描述划分结果与真实结构之间的相关性[15]，它的取值范围为[0.1]，值越接近 1 表示划分结果越好。数学描述如下：

$$NMI=1-\frac{1}{2}[H(X\,|\,Y)+H(Y\,|\,X)] \tag{10}$$

$$H(X\,|\,Y)=\frac{1}{|c_i|}\sum_k\frac{H(X_k\,|\,Y)}{H(X_k)} \tag{11}$$

其中：$X(Y)$ 是指划分社区 $c_i(c_j)$ 相关的随机变量，$H(X\,|\,Y)$ 是在划分 Y 下 X 的归一条件熵。

3.2　算法比较结果评价

模拟数据集主要由 LFR[16]基准图组成，选择不同的参数生成不同的 LFR 基准图，不同的参数生成不同类型的复杂网络。N 表示网络节点数目，$\bar{k}$ 为网络中节点的平均度数，$k_{\max}$ 是指网络中最大度数，u 表示可调的模糊参数（混合参数），O_n 与 O_m 分别是指重叠节点数目和重叠节点社区从属数目，τ_1 表示节点度的幂率分布系数，τ_2 为社区规模的幂率分布系统，$c_{\min}$ 与 $c_{\max}$ 分别是指网络中社区的最小规模和最大规模。当 LFR 基准网络参数设定为：$N=5000,\tau_1=2,\tau_2=1,\bar{k}=10,k_{\max}=50$，设置混合参数为 $\mu=0.3$，它是指在同一个社区内一个节点和其他节点的链接率；网络中社区规模设置为 $(20,100)$；重叠节点个数 O_n 设置为占网络所有节点的 10%；重复节点所属社区的个数 O_m 设置为 $2,3,4,5,6$；O_m 值越大，划分重叠社区结构的难度越大。而真实数据集的基本信息如表 1 所示。

表 1　七个真实数据集的基本信息[16]

Network	V	E	C
karate	34	78	2
dolphin	62	159	2
football	115	616	12
politics	105	613	3
Jazz	198	2743	Unknown
e-mail	1133	5451	Unknown
PGP	10680	24316	Unknown

三种算法在模拟数据集的 NMI 值受混合参数 μ 与重叠节点社区从属数目 O_m 的影响，随着它们越大，划分重叠社区结构的难度越大。图 2 所示，当 $\mu=0.3$ 时，随着 O_m 从 2~6 变化，三种算法的性能都会下降，与两种算法相比，LEGAOCD 算法仍能得到较高 NMI 值，说明本文所提的算法在划分模拟数据集网络中有了较好的性能。

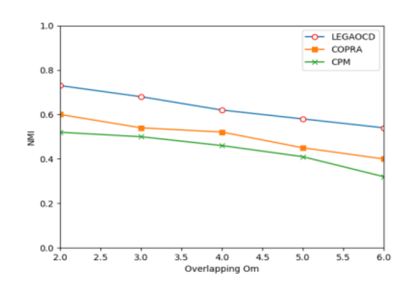


图 2　算法在模拟数据集上的 NMI 值对比结果

算法检测重叠社区结构的能力与检测重叠节点的能力不一定保持一致[18]。F-score 值比较结果如图 3 所示，LEGAOCD 算法会有高召回率和低准确率是由于它的后处理机制会检测出超过真实重叠的检测数目，所以，算法在检测重叠节点上性能表现一般，O_m 值越大，LEGAOCD 算法检测重叠节点的能力更好。

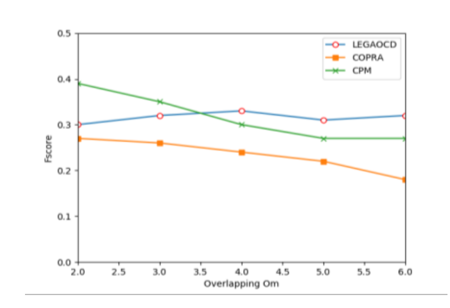


图 3　算法在模拟数据集上得到的 F 值对比结果

三种算法在真实数据集网络得到的 SQ 值都在 0.5 以上，具有较清楚的网络划分社区结构，如图 4 展示了 LEGAOCD 算法在真实数据集网络得到的 SQ 值都比其他算法较高（除 PGP 外），在较大的 PGP 网络中，LEGAOCD 算法比 CPM 算法的 SQ 值较高，但比 COPRA 算法的 SQ 值较低。

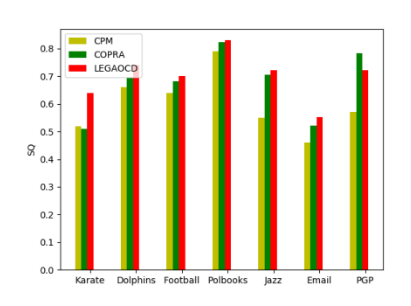


图 4　算法在真实数据集网络上得到的 SQ 值

4 结束语

本文提出局部扩展的遗传优化重叠社区发现算法,主要贡献体现在三方面：扩展的遗传算法以解决重叠社区发现问题；定义了模体稳定性的度量，从而量化社区稳定性；节点与社区的贴近度公式为遗传算法的适应度函数。面对大型网络，**LEGAOCD** 算法与 **CPM** 算法运行的时间和空间开销过高，如何在大型网络挖掘出较清楚的社团结构将是下一个研究的方向。

参考文献：

[1] 周旭. 复杂网络中社区发现算法研究 [D]. 长春: 吉林大学, 2016.

[2] Mahmoud H, Masulli F, Rovetta S, et al. Community detection in protein-protein interaction networks using spectral and graph approaches [C]// Computational Intelligence Methods for Bioinformatics, and Biostatistics. [S. l.] : Springer International Publishing, 2013: 62-75.

[3] Palla G, Derényi I, Farkas I. Uncovering the overlapping community structure of complex networks in nature and society. [J]. Nature, 2005, 435 (7043): 814-8.

[4] Gregory S. Finding overlapping communities in networks by label propagation [J]. New Journal of Physics, 2009, 12 (10): 2011-2024.

[5] Xie J, Szymanski B K, Liu X. SLPA: Uncovering Overlapping Communities in Social Networks via a Speaker-Listener Interaction Dynamic Process [J]. 2011: 344-349.

[6] 乔少杰, 郭俊, 韩楠, 等. 大规模复杂网络社区并行发现算法 [J]. 计算机学报, 2017, 40 (3): 687-700.

[7] 韩忠明, 谭旭升, 陈炎, 等. NCSS: 一种快速有效的复杂网络社团划分算法 [J]. 中国科学: 信息科学, 2016 (4) .

[8] Prat-Pérez A, Dominguez-Sal D, Brunat J M. Shaping Communities out of Triangles [J]. Computer Science, 2012: 1677-1681.

[9] 张海燕, 梁循, 周小平. 针对有向图的局部扩展的重叠社区发现算法 [J]. 数据采集与处理, 2015 (3): 683-693.

[10] 乔少杰, 韩楠, 张凯峰, 等. 复杂网络大数据中重叠社区检测算法 [J]. 软件学报, 2017, 28 (3): 631-647.

[11] Atay Y, Kodaz H. A New Adaptive Genetic Algorithm for Community Structure Detection [M]// Intelligent and Evolutionary Systems. [S. l.] : Springer International Publishing, 2016.

[12] 王琦, 温志平. 一种基于多维遗传算法的重叠社区发现方法 [J]. 计算机应用研究, 2016, 33 (12): 3543-3546.

[13] 牛新征, 司伟钰, 佘堃. 基于进化聚类的动态网络社团发现 [J]. 软件学报, 2017, 28 (7): 1773-1789.

[14] Shen B, Wang N, Qiu H. A new genetic algorithm for overlapping community detection [C]// Proc of the 10th International Conference on Intelligent Information Hiding and Multimedia Signal Processing. Washington DC: IEEE Computer Society, 2014: 766-769.

[15] Li Z, Liu J. A multi-agent genetic algorithm for community detection in complex networks [J]. Physica A Statistical Mechanics & Its Applications, 2016, 449: 336-347.

[16] Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms [J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2008, 78 (2): 046110.

[17] Zhou X, Liu Y, Zhang J, et al. An ant colony based algorithm for overlapping community detection in complex networks [J]. Physica A Statistical Mechanics & Its Applications, 2015, 427: 289-301.

[18] Xie J, kelley S, Szymanski, B. Overlapping community detection in networks: the state of the art and comparative study [J] , ACM Computing Surveys, 2013, 45 (4): 1-37.